Supplemental Online Content

Methods (extended)

Figure S1. Barcelona city map distribution of study's high schools.

Figure S2. Red blood cell alpha-linolenic acid (ALA) status before and after the intervention.

Table S1. Changes in secondary endpoints comparing the walnut group to the control group at 6 months using intention-to-treat analysis.

Table S2. Changes in primary endpoints comparing the walnut group to the control group at 6 months using intention-to-treat analysis adjusting for baseline outcomes values.

Table S3. Changes in secondary endpoints comparing the walnut group to the control group at 6 months using intention-to-treat analysis adjusting for baseline outcomes values.

Table S4. Per-protocol analyses corrected p-values for multiple testing using the Benjamini-Hochberg false discovery rate.

Methods (extended)

A protocol paper with extended description of the study design is already published.[1]

Neuropsychological and behavioural testing for primary endpoints (extended)
The first primary outcome was the Attention Network Test (ANT), which is a computerized test that examined sustained attention based on reaction times (RT) to target stimuli.[2] The test consisted of the appearance of five arrows on the screen and participants had to indicate the direction of the central arrow as fast as possible. The main ANT outcome selected was hit reaction time standard error (HRT-SE), which measures standard error of RT for correct responses, being an indicator of selective and sustained attention. The outcome is measured in milliseconds and lower scores indicate better attention performance.

The second measurement was the N-back task, which is also a computer-based test assessing working memory (WM) using series of numbers.[3] In this test, a sequence of items is shown, for each item it must be determined whether the last one shown is identical to the stimulus presented "n" trials back (1-, 2-, 3- and 4-back). These different conditions are known as loads and in the highest cognitive load (i.e., 4-back) the demands on WM are stronger. For this study, the 4-back test was used. In the 4-back level, the target was any stimulus identical to the one presented four trials previously. The more correct answers the higher the score. The outcome selected for this test was d prime (d'). Measures of d' were computed as follows: $d' = z(\text{hit rate}) - z(\text{false alarm rate})$. A higher d' indicates better detection, and thus, a more accurate performance.

The third measurement was an inductive reasoning subtest (i.e., fluid intelligence) of the Tests of Primary Mental Abilities (PMA-R, the Spanish adaptation).[4] The test consists of choosing a letter from a set of six possible alternatives, underlying a given sequence of letters. The total score is the number of correct item responses.

The fourth measurement was The Roulettes Task, a gambling task adapted from the Cups Task which assesses risky decision-making, i.e., whether the participants adjust their risky behaviours according to the probabilities and importance of the outcome.[5] The Roulettes Task is a game of chance whose goal is to make as much money as possible. Two roulettes appear on the screen, divided into segments. In each segment there is an amount of money. The left roulette has one segment with a larger amount of money (in this case $2), and the other segments are empty ($0). On the other hand, all the segments in the right roulette have the same amount ($1). The left roulette is a "risky" option since you don't know to which segment the roulette will stop turning. The more segments there are, the less likely it is to stop at a higher amount of money. The right-hand roulette is a "safe" option because it is certain that the roulette will stop on a $1 segment. Additionally, roulettes can be red or blue. The red ones lead to money loss and the blue ones lead to money gain. Participants must choose which roulette wheel they want to spin (for several times) and see if they adjust their risky behaviour according to the probabilities and importance of the outcome. The scores are measured as total risk adjustment to assess risk sensitivity. The closer to 0 the risk adjustment index is, the more risk insensitivity.

The fifth measurement was a self-reported version of the Strengths and Difficulties Questionnaire (SDQ, the Spanish adaptation), which gave us a total score of problem behavior.[6] It consists of 25 questions, organized in a general score of problem behaviour and five sub-scales aimed to assess emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and pro-social behaviour. A higher score indicates a more problematic behaviour.

The sixth measurement was the attention deficit hyperactivity disorder (ADHD) DSM-IV form list. We asked the teacher to answer 18 multiple-choice questions in order to rate the ADHD performance in a general score of the participant (adolescent). The scores were transformed into number of symptoms and diagnostic criteria for ADHD; A higher score indicates more ADHD related symptoms.[7]

Secondary endpoints (extended)
Height, weight, and waist circumference were measured by standard methods. Height was measured using the stadiometer model SECA 214, weight by the weighting scale SECA 770 model, and waist circumference with the SECA 201 tape model. Body mass index was calculated based on $\text{weight}/\text{height}^2$. All measurements were collected by a trained nurse at the schools at baseline and after 6 months of intervention.

For the determination of omega 3 fatty acids, a nurse obtained blood samples after an overnight fast. The samples were centrifuged at 2500g for 20 minutes at 20ºC within 4 hours after extraction. We obtained a 1500-μL aliquot of packed red blood cells (RBCs), which was stored at -80ºC until fatty acid analysis. We determined the fatty acid profile in RBC by gas-chromatography, as described.[8] In brief, cells contained in a 100-μL aliquot of RBCs were haemolysed and spun. The pellet membranes were dissolved in 1 mL BF_3 methanol solution and heated to hydrolyse and methylate glycerophospholipid fatty acids contained in the RBC membrane. The fatty acid methyl

esters were isolated by adding n-hexane and were separated by gas chromatography using an Agilent HP 7890 Gas Chromatograph equipped with a 30 m × 0·25 μm × 0·25 mm SupraWAX-280 capillary column (Teknokroma, Barcelona, Spain), an autosampler, and a flame ionization detector. The amount of each fatty acid was expressed as a percentage of the total identified fatty acids in the sample. The omega-3 fatty acids included in this study are ALA, DHA, and EPA.

Sample size calculation
The power of the study for neuropsychological outcomes was based on 400 participants per group. The six primary outcomes were considered from the six tests (N-back task, ANT, PMA-R, The Roulettes Task, SDQ, and the ADHD DSM-IV form list). The cognitive outcomes with mean values of 100 (SD 15) units corresponding to the standard values of neuropsychological scores in the general population were considered, with a correlation between them of 0·25. The targeted intervention effect was a change of 3 units. We considered a type-I error of 0·05 and corrected calculations for multiplicity using the Benjamini-Hochberg method. A 10% loss of follow-up was assumed. Additionally, we assumed that the final models had an R^2 of 20%. With all these considerations, the study would have 95% power to detect the association with at least one outcome, 90% to detect at least two, 80% to detect at least 3, 70% to detect at least 4, 55% to detect at least 5, and 31% to detect the effect with all six outcomes.

References

1   Julvez J, Gignac F, Fernández-Barrés S, *et al.* Walnuts, Long-Chain Polyunsaturated Fatty Acids, and Adolescent Brain Development: Protocol for the Walnuts Smart Snack Dietary Intervention Trial. *Front Pediatr* 2021; **9**: 425.

2   Forns J, Esnaola M, López-Vicente M, *et al.* The n-back test and the attentional network task as measures of child neuropsychological development in epidemiological studies. *Neuropsychology* 2014; **28**: 519–29.

3   López-Vicente M, Forns J, Suades-González E, *et al.* Developmental trajectories in primary schoolchildren using n-back task. *Front Psychol* 2016; **7**. DOI:10.3389/FPSYG.2016.00716.

4   Thurstone TG. The Tests of Primary Mental Abilities. *Pers Guid J* 1957; **35**: 569–76.

5   Levin IP, Weller JA, Pederson AA, Harshman LA. Age-related differences in adaptive decision making: Sensitivity to expected value in risky choice. *Judgm Decis Mak* 2007; **2**: 225–33.

6   Ortuño-Sierra J, Fonseca-Pedrero E, Paino M, Sastre I Riba S, Muñiz J. Screening mental health problems during adolescence: psychometric properties of the Spanish version of the Strengths and Difficulties Questionnaire. *J Adolesc* 2015; **38**: 49–56.

7   American Psychiatric Association.2000. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR® (4th ed, text revision). Washington, DCAmerican Psychiatric Association.

8   Sala-Vila A, Harris WS, Cofán M, *et al.* Determinants of the omega-3 index in a Mediterranean population at increased risk for CHD. *Br J Nutr* 2011; **106**: 425–31.

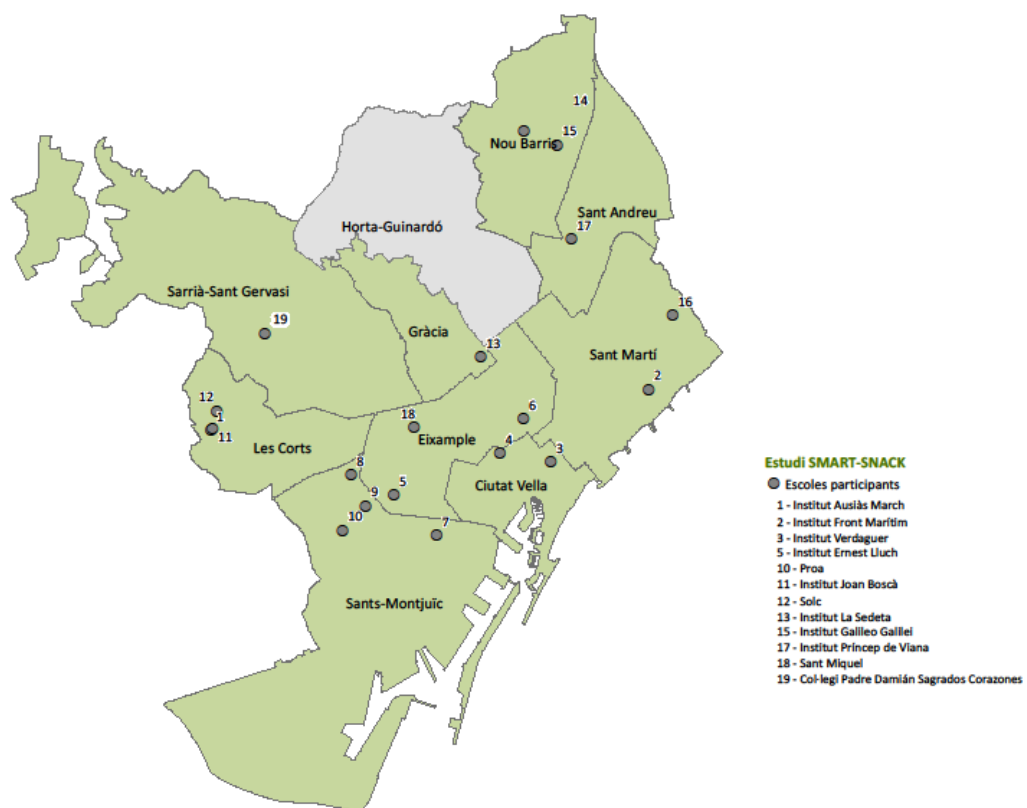**Figure S1.** Barcelona city map distribution of study's high schools.

Estudi SMART-SNACK

○ Escoles participants

1 - Institut Ausiàs March
2 - Institut Front Marítim
3 - Institut Verdaguer
5 - Institut Ernest Lluch
10 - Proa
11 - Institut Joan Boscà
12 - Solc
13 - Institut La Sedeta
15 - Institut Galileo Galilei
17 - Institut Príncep de Viana
18 - Sant Miquel
19 - Col·legi Padre Damián Sagrados Corazones

Figure S2. Red blood cell alpha-linolenic acid (ALA) status before and after the intervention.
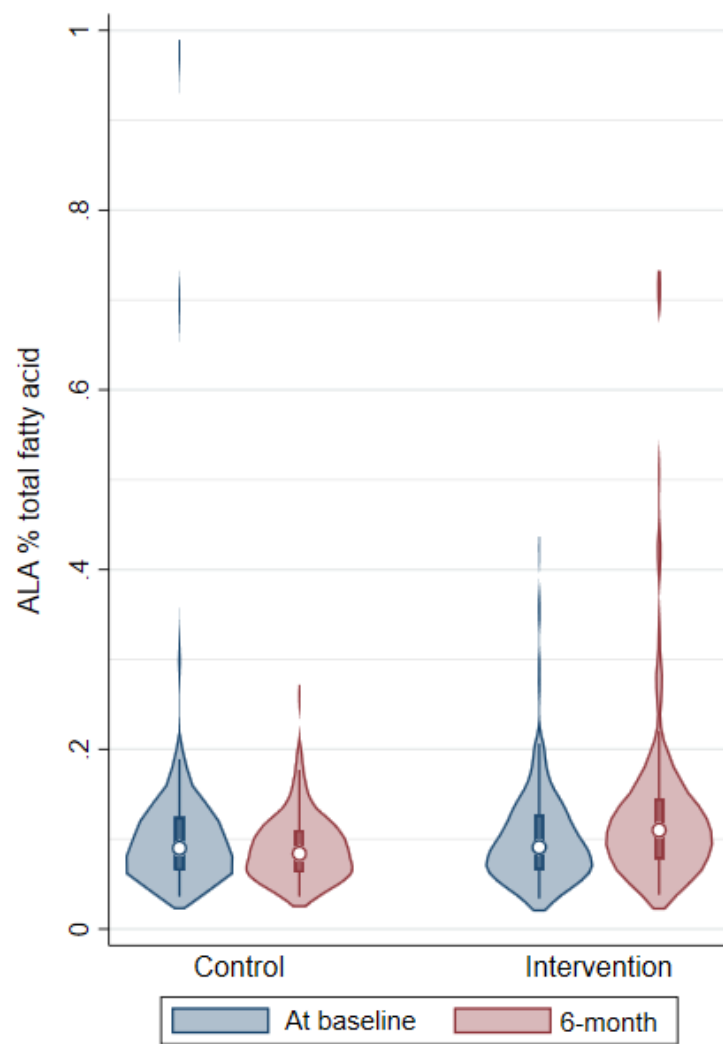
Table S1. Changes in secondary endpoints comparing the walnut group to the control group at 6 months using intention-to-treat analysis.

Outcomes after 6-month intervention	N	β Coef.[a]	95% CI	*p*-Value
Height (m)	644	0·00	(-0·01, 0·01)	0·59
Weight (kg)	644	0·19	(-1·44, 1·81)	0·82
Waist circumference (cm)	282	0·20	(-1·58, 1·99)	0·82
BMI	644	0·01	(-0·47, 0·49)	0·97
ALA (%)	270	0·04	(0·03, 0·06)	< 0·0001
EPA (%)	270	0·02	(-0·02, 0·05)	0·34
DHA (%)	270	-0·05	(-0·24, 0·14)	0·60

N, number of observations for each analysis.

[a] Data were calculated using a linear mixed model with school and household as random effects and adjusted for sex, age, and maternal education.

Table S2. Changes in primary endpoints comparing the walnut group to the control group at 6 months using intention-to-treat analysis adjusting for baseline outcomes values.

Outcomes after 6-month intervention	N	β Coef.[a]	95% CI	*p*-Value
Attention score (ms)	640	0·07	(-9·30, 9·44)	0·99
Working memory score	627	-0·03	(-0·21, 0·15)	0·75
Fluid intelligence score	638	-0·11	(-0·64, 0·41)	0·67
Risky decision-making score	640	-0·12	(-0·67, 0·44)	0·68
Behavioural problem score	331	0·72	(-0·28, 1·72)	0·16
ADHD general score	509	0·15	(-1·03, 1·34)	0·80

N, number of observations for each analysis.

[a] Data were calculated using a linear mixed model with school and household as random effects and adjusted for sex, age and maternal education, and further adjusted for the baseline outcome variable.

Table S3. Changes in secondary endpoints comparing the walnut group to the control group at 6 months using intention-to-treat analysis adjusting for baseline outcomes values.

Outcomes after 6-month intervention	N	β Coef.[a]	95% CI	*p*-Value
Height (m)	638	0·00	(-0·00, 0·00)	0·78
Weight (kg)	638	-0·13	(-0·60, 0·34)	0·60
Waist circumference (cm)	277	0·13	(-0·56, 0·81)	0·72
BMI	638	-0·05	(-0·22, 0·12)	0·54
ALA (%)	265	0·04	(0·03, 0·06)	< 0·0001
EPA (%)	265	0·02	(-0·01, 0·05)	0·12
DHA (%)	265	-0·09	(-0·21, 0·03)	0·13

N, number of observations for each analysis.

[a] Data were calculated using a linear mixed model with school and household as random effects and adjusted for sex, age and maternal education, and further adjusted for the baseline outcome variable.

Table S4. Per-protocol analyses (Table 4) corrected p-values for multiple testing using the Benjamini-Hochberg false discovery rate.

Outcomes after 6-month intervention	p-value	Benjamini-Hochberg critical value[a]
Fluid intelligence score[*]	<0·001	0·008
ADHD general score[*]	0·005	0·017
Attention score (ms)[*]	0·011	0·025
Risky decision-making score	0·382	0·033
Behavioural problem score	0·587	0·042
Working memory score	0·836	0·050

[a] False Discovery Rate (FDR) significance threshold was defined at 0.05. Tests are considered statistically significant if their p-values are smaller than the largest p-value that is less than its Benjamini-Hochberg critical value (*).